Online Materials and Methods

EudraCT version

All data analyzed for our study resulted from queries performed from the EudraCT database, current Version 8.1.1. Version 8 of EudraCT was released in March 2011 and involved a major change to the EudraCT system and how the data are stored. These changes were made in preparation for the launch of the EU Clinical Trials Register in March 2011, which is a public database that displays information from within EudraCT; important data such as phase I clinical trials in adults are, however, not publicly available. This article contains data up to 31.12.2010. EudraCT only contains data from trials with sites in the EU/EEA and therefore does not encompass data on trials conducted wholly outside the European Economic Area. Such clinical trials have from March 2011 onwards become part of EudraCT but only when they are part of a Paediatric Investigational Plan. Therefore, our study gives a more precise picture of the situation in Europe. Each clinical trial with at least one clinical investigator site in Europe receives a unique number for identification, the 'EudraCT Number'. The EudraCT Number must be included on all Clinical Trial applications within Europe and as needed on other documents relating to the trials (e.g. Suspected Unexpected Serious Adverse Reaction (SUSAR) reports, for which reporting to regulators by sponsors is mandatory for clinical trials in Europe). It represents a unique key that allows identifying a clinical trial and its details, and enabled the generation of the data discussed in this article.

Generation of raw data from EudraCT

Specific queries were designed to extract data from EudraCT concerning all clinical trials on ATMPs commencing from 1 May 2004 to 31 December 2010, i.e. spanning the time frame from the very first initiation of the EudraCT register until the time the Committee for Advanced Therapies entered the last third of its first term (i.e., January 2011), which the committee considered a milestone. Queries were entered to extract the name of the product, sponsor, indication, member states where the trial was

performed, and others (the following EudraCT fields were included in the queries: A.1 Application NCA, A.2 EudraCT number, CTA Load Year, A.3 Full title of the trial, B.1.1 Sponsor Organisation, B.1.2 Sponsor Contact Given name, B.1.2 Sponsor Contact Family Name, B.1.3 Sponsor Street Address, B.1.3 Sponsor Town/City, B.1.3 Sponsor Post Code, B.1.3 Sponsor Country, B.1.4 Sponsor Telephone, B.1.6 Sponsor Email, B.3.1 and B.3.2 Sponsor Status, D.1.2 AND D.1.3 IMP Category, D.2.1.1.1 IMP Trade name, D.2.5 IMP. is orphan drug, D.2.5.1 Orphan drug number, D.3.1 IMP Name, D.3.2 IMP Code, D.3.3 IMP ATC Code, D.3.8 AS INN, D.3.11.3 Somatic cell therapy MP, D.3.11.4 Gene therapy MP, E.1.1 Medical condition, E.2.1 Trial main objective, E.2.2 Trial secondary objective, E.7.1 Trial type Human pharmacology (Phase I), E.7.2 Trial type Therapeutic Exploratory (Phase II), E.7.3 Trial type Therapeutic Confirmatory (Phase III), E.7.4 Trial type Therapeutic Use (Phase IV), F.1.1 Population under eighteen, P. MS Trial End, P. MS Trial End Date, P. Global Trial End, P. Global Trial End date, P. EOT type, Status of the Trial).

Initially these queries resulted in a Microsoft Excel report extracted from EudraCT containing 642 records (i.e. lines in the Excel report). These records included multiple entries for the same clinical trial (corresponding to more than one record with the same EudraCT number). This is explained by the fact that when a clinical trial is performed in more than one Member State, each Member State enters one set of data (record) in EudraCT by loading the CTAs received by the sponsor of the trial. Also, when the clinical trial (CT) is performed with more than one substance, e.g. the medicine and a comparator, a double entry (i.e. line) appears in the report extracted by EudraCT.

Data cleaning

In order to eliminate these multiple records but retain the information originally entered in EudraCT, the data were uploaded in a specifically designed FileMaker database (designed by L. D'Apote and C. Pereira, European Medicines Agency) (FileMaker Pro Advanced 9.0v3, FileMaker Inc.), a relational

database specifically designed to import and clean the dataset generated by EudraCT. It was capable to detect and highlight all records with the same EudraCT number, allowing the operator to label a record as 'unique', 'master' or 'copy'. Records marked as 'unique' correspond to clinical trials appeared only once in EudraCT. Multiple records with the same EudraCT number were marked once as 'master', and the remaining records were marked as 'copy'. The latter did not appear in the filtered record list unless specifically requested by the operator. The filtered record list showed only one record for each clinical trial.

Records that concerned products which were wrongly categorized and entered as ATMPs in the EudraCT base by the sponsor were reviewed in order to confirm that these belong to other categories of products that were not formally ATMPs.

We verified the given classification based on the description of the products, supplemented, where necessary, by literature search on invented names or abbreviations entered by the sponsor. Where needed, we re-classified products into the respective ATMP category (gene therapy medicinal product or cell-based medicinal product including one of its subclasses, e.g. tissue engineered product).

Searches were independently performed in double by two operators. These two sets of results were compared, and discrepancies were re-analysed by check of the original dataset.

Data mining

Further screening of the cleaned data was performed using the specifically designed FileMaker database as described above in order to obtain stratification of clinical trials per nationality of the sponsor (using the search field 'sponsor country'), followed by use of further filters by year of start of the clinical trial and per clinical development phase (phase I, phase II, phase III, post-authorisation, which corresponds to phase IV). Using semi-automated searches, the operators filtered the clinical trials as national clinical trials, i.e. performed in only a single member state and multi-national clinical trials, i.e. performed in at

least two member states. The method allowed determining the exact number of sponsors. Some sponsors performed more than one clinical trial and were hence counted only once. The data were further analyzed by EudraCT field 'Sponsor Status' (fields B.3.1 and B.3.2 of the clinical trials loaded in EudraCT, see Ref. 22) in two major categories of as 'commercial' and 'non-commercial'. (N.B.: In 4 cases the field concerning the details of the sponsor and their commercial status was left empty in the respective EudraCT entry, which led to a total number of clinical trials performed by commercial and non commercial sponsors being 314 rather than 318).

Subsequently the following analyses were performed manually by the operators: commercial sponsors were further classified in three categories: 'Large pharmaceutical companies', 'SME' and 'non registered SME/not large pharmaceutical company'. To determine whether the commercial sponsors were registered as SMEs, the list of sponsors was compared with the European Medicines Agency public register of SMEs, which includes companies established in the European Economic Area that have submitted a SME declaration within the scope of Recommendation 2003/361/EC (7) and to whom the Agency has assigned SME status (8). Sponsors were classified as 'large pharmaceutical companies' from information sourced from their web sites when it was clear that they employ more than 250 persons and have either an annual turnover exceeding 50 million Euro, or an annual balance sheet total exceeding 43 million Euro (which are the criteria for defining an enterprise as SMEs, see European Commission (2011): The new SME definition. User guide and model declaration. ISBN 92-894-7909-4). Sponsors which were not included in the EMA public register of SMEs or from the information provided on their website could not be classified as 'large pharmaceutical companies' were included in the category 'non registered SME/non big pharma'. Non-commercial sponsors were also divided into 'academia' or 'charity' based on the information entered in EudraCT (affiliation given as 'university' or 'Trust'). In case of unclear status of a sponsor, its category was determined by on-line literature searches.

The category of cell-based medicinal products was further stratified into dendritic cells, mesenchymal stem cells, haematopoietic stem cells or bone marrow-derived stem cells in homologous use, but substantially manipulated, haematopoietic stem cells or bone marrow stem cells in non-homologous (heterologous) use, tissue engineered products, and 'Other cell based medicinal products'.

Of note, "homologous use" means that a cell-based medicinal product is used for the same essential function, e.g. classical bone marrow transplantation. Such products are in Europe not included in the ATMP legislation, but they would qualify as ATMP if they are substantially manipulated. Substantial manipulation is, for example, cell culture expansion or differentiation with cytokines or growth factors, while certain more simple manipulations like cutting, grinding, shaping, centrifugation, or sterilization and others would not be considered substantial manipulation (see Annex I to Regulation 1394/2007, Ref. 2). Products that fall under this category of are, for example, allogeneic T cells encoding a thymidine kinase gene, activated natural killer cells, or substantially modified (like cytokine-activated) T lymphocytes (see also the published summaries of scientific recommendations on classification of ATMPs issued by the CAT, Ref. 9). Haematopoietic stem cells or bone marrow-derived stem cells in non-homologous use are defined by the ATMP Regulation as "engineered" and are therefore ATMPs (Regulation No 1394/2007, Article 2, 1.c: *"The cells or tissues are not intended to be used for the same essential function or functions in the recipient as in the donor"*, Ref. 2). For example, minimally manipulated bone marrow-derived stem cells for the treatment of myocardial infarction fall under this category. Taken together, the domain of stem cell-based medicinal products (i.e., MSC plus haematopoietic stem cells or bone marrow-derived stem cells in homologous and non-homologous use) was around half (49%) of the cell-based medicinal products studied. For analysis of products in studied in solid tumours, products were counted only once if they were studied in more than one trial in a particular indication (e.g., phase I/II study followed by a phase III study within the observation period) in order to avoid double counting. If a given product was studied in more than one oncological indication,

then these were counted separately (since every indication is an own stand-alone development for a product).

Data concerning orphan drugs were extracted using specific EudraCT fields as filters (fields D.2.5 ‘IMP is orphan drug’ and D.2.5.1 ‘Orphan drug number’ of the Clinical Trial Authorisation form).